and H32. The NOEs clearly define the orientation of the ligand peptide on the WW domain.

The proline-containing part of peptide GTPPPPYTVG was restrained into a polyproline helix type II, and the structure of the complex was calculated using a standard simulated annealing protocol[15] (X-PLOR[16]) and the ten intermolecular NOEs as constraints. The result shows smooth contacts between the ligand and the domain (Fig. 3). The central prolines P4′ and P5′ contact W39, and the carbonyl group of P6′ points towards the OH group of the conserved residue Y28. The peptide tyrosyl residue Y7′ is accommodated in a hydrophobic pocket formed by L30 and H32. These contacts are well defined by six NOEs between the aromatic ring of Y7′ and the side chains of L30 and H32. Y7′ could form a hydrogen bond to the histidine ring but also to Q35, whose chemical shifts change strongly on peptide binding (Fig. 1).

The aromatic residues at positions 39 and 28, a hydrophobic residue at position 30 and a histidine at position 32 all tend to be conserved (Fig. 1). Our structure shows that these are the residues that are in contact with the peptide. The importance of this hydrophobic surface is underscored by the low binding affinity of mutants H32A, L30K and Q35A to both peptides (except in one case, see Table 1; E.B. *et al.*, unpublished results). The structure of the mutants remained intact, as judged by their two-dimensional NMR spectra. The binding affinities of peptides in which the tyrosine residue of SPPPPYTV is substituted with alanine, leucine or phenylalanine show that the YAP domain is specific for a tyrosine-containing motif. In the first two cases there is no binding, and the peptide containing phenylalanine has only a weak affinity (Table 1). Thus, the tyrosine in the PPxY motif may be needed for the interaction of ligands with a set of WW domains. The importance of the polyproline motif is shown by the lack of binding when the two centre prolines in this peptide are replaced by alanines, whereas replacement of the first proline has no effect (Table 1).

Our structure confirms the hypothesis[4] that the WW domain is a binding module for proline-rich ligands. The PPxY motif may not be the only ligand for WW domains: for instance, it is not present in the proline-rich tail of formin that interacts with SH3 and WW domains[10], where sequence motifs such as PPxLP are found. The structure indicates that hydrophobic residues replacing tyrosine of the ligand could be accommodated on the surface of those domains that contain hydrophobic residues other than leucine at position 30.

Note added in proof: An involvement of hYAP in retroviral budding through its WW domain was recently suggested[27]. □

Received 17 May; accepted 24 June 1996.

1. Bork, P. & Sudol, M. *Trends biochem. Sci.* **19,** 531–533 (1994).
2. Andre, B. & Springael, J. Y. *Biochem. biophys. Res. Commun.* **205,** 1201–1205 (1994).
3. Hofmann, K. & Bucher, P. *FEBS Lett.* **358,** 153–157 (1995).
4. Chen, H. I. & Sudol, M. *Proc. natn. Acad. Sci. U.S.A.* **92,** 7819–7823 (1995).
5. Pawson, T. *Nature* **373,** 573–580 (1995).
6. Cohen, G. B., Ren, R. & Baltimore, D. *Cell* **80,** 237–248 (1995).
7. Sudol, M., Chen, H. I., Bougeret, C., Einbond, A. & Bork, P. *FEBS Lett.* **369,** 67–71 (1995).
8. Bork, P. & Margolis, B. *Cell* **80,** 693–694 (1995).
9. Zhou, M. M. *et al. Nature* **378,** 587–592 (1995).
10. Chan, D. C., Bedford, M. T. & Leder, P. *EMBO J.* **15,** 1045–1054 (1996).
11. Cicchetti, P., Mayer, B. J., Thiel, G. & Baltimore, D. *Science* **257,** 803–806 (1992).
12. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. *Science* **259,** 1157–1161 (1993).
13. Staub, O. *et al. EMBO J.* **15,** 2371–2380 (1996).
14. Schild, L. *et al. EMBO J.* **15,** 2381–2387 (1996).
15. Nilges, M., Clore, G. M. & Gronenborn, A. M. *FEBS Lett.* **229,** 317–324 (1988).
16. Brünger, A. T. *X-PLOR 3.1 Manual* (Yale Univ. Press, New Haven, 1992).
17. Viguera, A. R., Arrondo, J. L. R., Musacchio, A., Saraste, M. & Serrano, L. *Biochemistry* **133,** 10925–10933 (1994).
18. Aue, W. P., Bartholdi, E. & Ernst, R. R. *J. chem. Phys.* **64,** 2229–2246 (1976).
19. Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. *J. Am. chem. Soc.* **110,** 7870–7871 (1988).
20. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. *J. chem. Phys.* **71,** 4546–4553 (1979).
21. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. *J. magn. Reson.* **89,** 496–514 (1990).
22. Fesik, S. W. & Zuiderweg, E. R. P. *J. magn. Reson.* **67,** 588–593 (1988).
23. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. *J. Am. chem. Soc.* **111,** 1515–1517 (1989).
24. Otting, G. & Wüthrich, K. *J. magn. Reson.* **85,** 586–594 (1989).
25. Kraulis, P. *J. appl. Crystallogr.* **24,** 946–950 (1991).
26. Nicholls, A., Sharp, K. A. & Honig, B. *Proteins* **11,** 281–296 (1991).
27. Garnier, L., Wills, J. W., Verderame, M. F. & Sudol, M. *Nature* **381,** 744–745 (1996).

ACKNOWLEDGEMENTS. We thank A. Ulrich for her participation in preliminary experiments; P. Bork, Henry I. Chen and M. Wilmanns for discussion; M. Wilmanns for help in preparing figures; R. Jacob for synthesis of peptides; A. Viguera for help with fluorescence measurements; and H. Kessler and C. Schwarz for facilities and help at the Technische Universität, München. M.S. was supported by MDA and HFSPO grants.

CORRESPONDENCE and requests for materials should be addressed to H.O. (e-mail: oschlinat@embl-heidelberg.de).

Crystal structure of a PDZ domain

João H. Morais Cabral*, Carlo Petosa*, Michael J. Sutcliffe†, Sami Raza†, Olwyn Byron*‖, Florence Poy‡, Shirin M. Marfatia§, Athar H. Chishti§ & Robert C. Liddington*

* Departments of Biochemistry, ‖ NCMH and † Chemistry, University of Leicester, Leicester LE1 7RH, UK
‡ Dana–Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA
§ Laboratory of Tumor Cell Biology, St Elizabeth's Medical Center, Tufts University School of Medicine, 736 Cambridge Street, Boston, Massachusetts 02135, USA

PDZ domains (also known as DHR domains or GLGF repeats) are ~90-residue repeats found in a number of proteins implicated in ion-channel and receptor clustering, and the linking of receptors to effector enzymes[1]. PDZ domains are protein-recognition modules; some recognize proteins containing the consensus carboxy-terminal tripeptide motif S/TXV with high specificity[2–4]. Other PDZ domains form homotypic dimers: the PDZ domain of the neuronal enzyme nitric oxide synthase binds to the PDZ domain of PSD-95, an interaction that has been implicated in its synaptic association[5]. Here we report the crystal structure of the third PDZ domain of the human homologue of the *Drosophila discs-large* tumour-suppressor gene product, DlgA. It consists of a five-stranded antiparallel β-barrel flanked by three α-helices. A groove runs over the surface of the domain, ending in a conserved hydrophobic pocket and a buried arginine; we suggest that this is the binding site for the C-terminal peptide.

The PDZ domain[6] is named after three of the proteins in which the repeats have been described: PSD-95 (postsynaptic density protein, M_r 95K), Dlg (*discs-large* protein) and ZO-1 (zonula occludens-1). These proteins have a conserved structure comprising three tandem PDZ domains, an SH3 domain and a guanylate-kinase-like domain. Other proteins that contain PDZ domains include certain protein kinases and protein tyrosine phosphatases, and neuronal nitric oxide (NO) synthase[1]. These domains appear to be protein-recognition modules analogous to the well characterized SH2 and SH3 domains.

We crystallized a recombinant form of the third PDZ domain (PDZ-3) from human Dlg and solved its structure at 2.8 Å resolution using two heavy-atom derivatives and solvent flattening (Table 1). A complete model for the 96-residue domain has been built; in spite of a large proportion of glycine residues (13%), all of the secondary structure elements and connecting loops are well ordered. The domain is compact and globular, with a diameter of 25–30 Å. β-strands 2–5 form an up-and-down β-barrel and strand β1 crosses over the barrel and hydrogen-bonds to β5; a short α-helix (α1) and its connecting loop cap one end of the barrel; helix α2 caps the other end of the barrel, and a C-terminal helix (α3) packs against the outside of the barrel (Figs. 1*b* and 2). Most of the conserved residues (Fig. 1*a*) are hydrophobic, and form the core of the domain, which is exposed on one face (Figs 2, 3*b*). There are two exceptions: a conserved aspartic acid (D510), which is buried and forms a salt bridge to an arginine (R465); and an asparagine

(N516) in the β4–α2 loop whose side chain packs against the α2–β5 loop. Insertions and deletions in other members of the PDZ family are restricted to the connecting loops between secondary structure elements. One major insertion of six residues occurs in PDZ-1 and PDZ-2 of hDLG and PSD-95, which maps to the loop between strands β2 and β3. A well-ordered helix, α3, extends seven residues beyond the C terminus of the published consensus sequence for the PDZ family, and may well be present in all members of the family. If this is the case, no or very few residues link the C-terminal helix of PDZ domain 1 (PDZ-1) to the first strand of PDZ-2, consistent with the finding that a recombinant fragment containing both PDZ-1 and PDZ-2 forms a single protease-resistant module (S.M.M. and A.H.C., manuscript in preparation).

Several PDZ domains (the closely related second domains of Dlg (ref. 4) and PSD-95 (refs 2, 3) and one domain of PTP-BAS (ref. 7)) recognize peptides containing a consensus C-terminal T/SXV sequence (in single letter amino-acid code, where X is any residue) with high specificity: mutation of the threonine to alanine or of the terminal valine to either alanine or aspartate abolishes binding[2]. A large number of membrane-associated proteins, including many neuronal ion channels and synaptic receptors, share this terminal motif[2,3]. We used the program SURFNET[8] to search for peptide-binding cavities on the surface of the domain, which uses only geometric criteria derived from the atomic coordinates. In known protein–ligand complexes, the program correctly predicts the ligand-binding site (corresponding to the largest cavity) for 85% of the cases (J. Thornton, personal communication). For the PDZ domain, the program finds only one cavity (volume, 720 Å^3) that is large enough for ligand binding (the volume of the second largest cavity is only 150 Å^3). The cavity includes a hydrophobic pocket and a groove that runs over the top of the molecule (Fig. 3*a*). The hydrophobic pocket is highly conserved among PDZ family members (Fig. 3*b*) and is formed by the β1–β2 loop (which includes the conserved GLGF motif) and side chains from strand β2 and helix α2 (Figs 2, 3*c*). At the

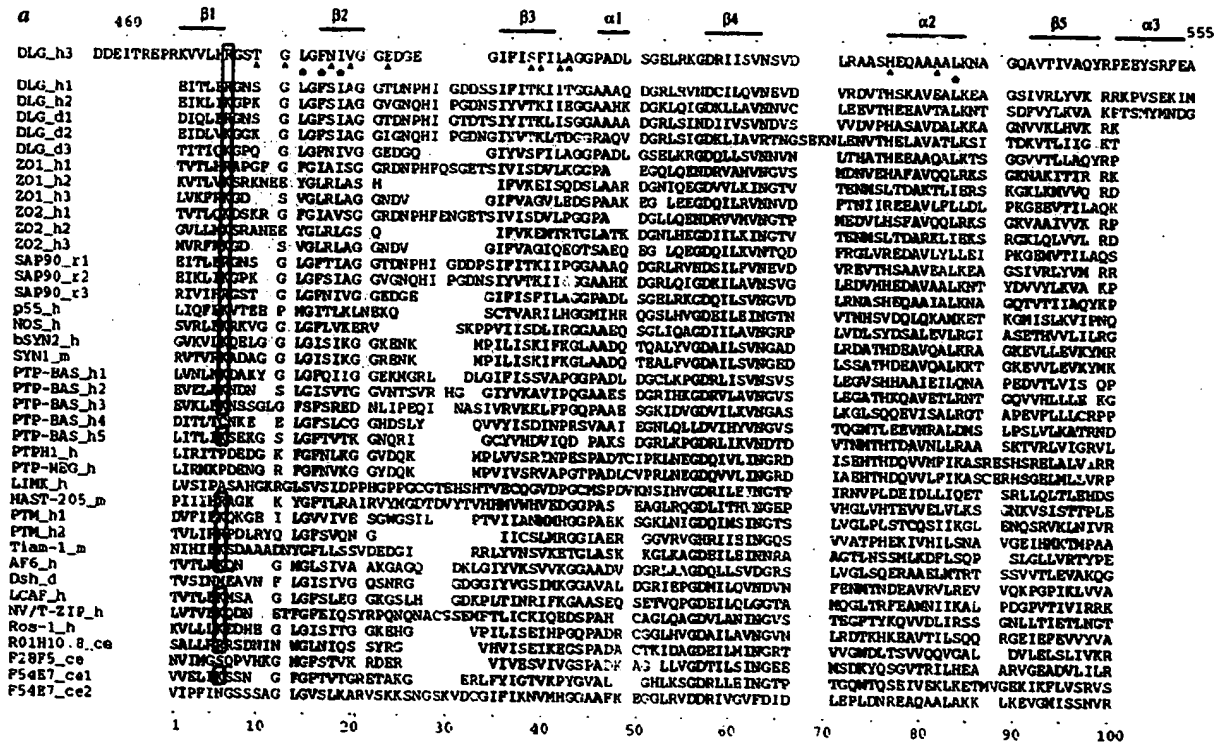
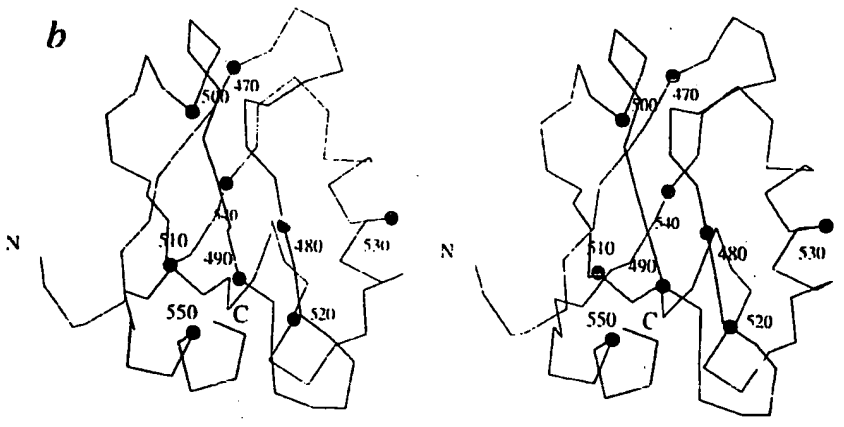
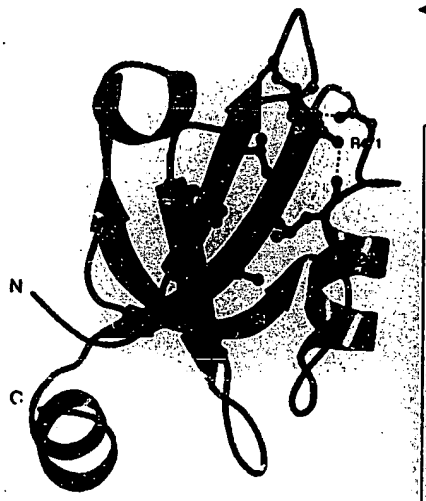


FIG. 1 *a*, Sequence alignment of the PDZ domain family, with secondary structure indicated for the PDZ-3 domain of hDlg (DLG_h3). Residue numbering at the top is that of the full-length hDlg protein. Positions where the chemical character of residues is conserved in 90% of sequences are highlighted in yellow. The four residues forming the hydrophobic pocket are marked by asterisks. The sequences of domains PDZ-1 and PDZ-2 from hDlg are contiguous. The sequences are from various species (h, human, r, rat; m, mouse; d, *Drosophila*; ce, *C. elegans*). The alignment and the sequence numbering shown at the bottom were also taken from ref. 1, where a full explanation of the sequence abbreviations can be found. The largely conserved basic residue corresponding to R471 in hDlg is boxed. The principal residues involved in dimerization are T474, G475, N479, V481, E484, S492, F493, L495, A496, H525 and A530 and are indicated with a caret (^). *b*, Stereo Cα plot of PDZ-3, with every tenth residue numbered and the N and C termini indicated. The fold-search program DALI[19] specifies the β-subunit of *Klebsiella aerogenes* urease[20] as having a similar fold, classified by the SCOP database[21] as a 'β-clip' fold, which is also shared by the enzyme dUTPase[22].





◀ FIG. 2 Ribbon diagram of the PDZ-3 domain with secondary structure elements and N and C termini indicated, generated with MOLSCRIPT[23], RASTER3D[24] and RENDER[25]. Side chains forming the hydrophobic pocket (residues Leu 476, Phe 478, Ile 480 and Leu 532), along with Arg 471, are also shown.

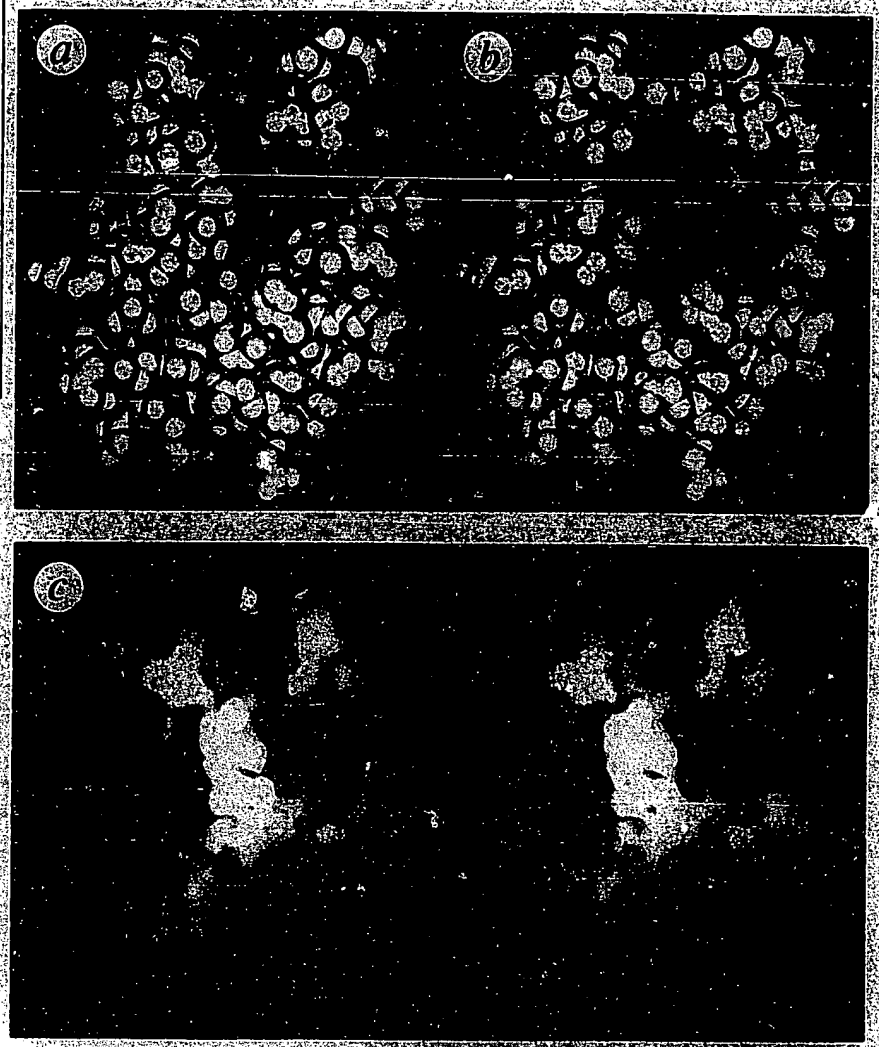


FIG. 3 Surface of the PDZ-3 domain. The view shown is similar to that in Fig. 2. *a*, Space-filling model with the largest cavity found by SURFNET[6], shown in gold. *b*, Space-filling model showing conserved residues. Residues in red are those that are highly conserved among PDZ domains. The side-chain nitrogens of Arg 471 are shown in blue. *c*, Stereo surface-charge representation, generated with the program GRASP[26]. Regions with positive and negative electrostatic potential are shown in blue and red, respectively. *d*, Close-up of *c*, with the model tripeptide TDV shown in the hydrophobic pocket.

METHODS. A selection of probes (-OH, CO_2^-, -CO, $-CH_3$, -NH) was used to search the SURFNET[6] cavity for possible binding sites using GRID[27]. The tripeptide was then manually docked into the cavity using the contour maps from GRID as a guide. Weak harmonic restraints were then applied to enable the automatic generation by molecular dynamics and simulated annealing (MODELLER[28]) of a set of 10 three-dimensional models of the tripeptide complexed with the PDZ domain. The model with the lowest energy was selected.

TABLE 1 Summary of crystallographic analysis

Crystals	Native	K_2PtCl_4	CH_3HgNO_3
Resolution (Å)	2.8	3.5	5.0
Unique reflections	5,793	1,998	1,014
Completeness (%)	100	80.1	98.6
R_{merge} (%) (outer shell)	10.8 (30.1)	8.6 (12.6)	4.0 (4.4)
Redundancy	13	2.5	9
R_{iso} (%)		9.6	5.6
No. of sites		3	2
Phasing power		1.5	0.62
Figure of merit, 0.34 (after DM[14] 0.76)			

Refinement statistics

Resolution	No. of atoms	R_{cryst} (%)	R_{free} (%)	R.m.s. bond lengths (Å)	R.m.s. bond angle (deg)
15–2.8 Å	727	22.0	27.1	0.008	1.4

$R_{merge} = \Sigma\Sigma_i|\langle I\rangle - I_i|/\Sigma\langle I\rangle$, where I_i is the observed intensity and $\langle I\rangle$ is the average intensity. $R_{iso} = \Sigma|F_{ph} - F_p|/\Sigma|F_p|$, where F_{ph} is the heavy-atom derivative structure factor and F_p is the protein structure factor. Phasing power is $\langle|F_{hc}|\rangle/\langle E\rangle$, where E is the residual lack of closure error. Figure of merit is $\langle|\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)|\rangle$, where $P(\alpha)$ is the phase probability distribution. $R_{cryst} = \Sigma|F_p - F_{p(calc)}|/\Sigma F_p$ for reflections with $F_p > 2\sigma F_p$. R_{free} is the same as R_{cryst}, but calculated on the 10% of data excluded from refinement. The domain was expressed as a glutathione-S-transferase (GST) fusion in *E. coli* strain BL21. The expressed protein includes residues 457 to 552 of the human homologue of *Drosophila discs-large* protein[11], and was purified on glutathione-Sepharose (Pharmacia) and eluted after protease digestion with thrombin. It was further purified on Mono-Q Sepharose (Pharmacia). The crystals grow by hanging-drop vapour diffusion at a protein concentration of 15 mg ml^{-1} from 0.9–1.3 M sodium citrate, pH 6.5–7.5 at 4 °C, and adopt space group $P6_322$, with $a = 111.0$ Å, $c = 62.5$ Å. All diffraction data were collected with a Rigaku RU200HB X-ray generator and image plate with Cu-Kα radiation. Data were processed using DENZO and SCALEPACK[12]. One platinum position was found from the difference Patterson map with the program RSPS[13], and further sites were derived from difference Fouriers after solvent flattening and histogram matching using the program DM[14] from the CCP4 program suite[15]. Heavy-atom parameters were refined using HEAVY[16] and MLPHARE[15]. The high solvent content (70%) present in the crystal was a powerful constraint during phase refinement and extension and allowed the production of an easily interpretable electron-density map[17]. The refinement consisted of rounds of simulated annealing and grouped *B*-factor refinement to 2.8 Å with XPLOR[18]. The present model includes 96 residues, from residues 460 to 552, plus three residues at the C terminus from the expression vector, and eight water molecules. All main-chain torsion angles, except S517, in the β4–α2 loop, fall in allowed regions of the Ramachandran plot. The relative molecular mass and K_d were determined by sedimentation equilibrium in a Beckman Optima XL-A analytical ultracentrifuge using a Marquardt–Levenberg non-linear least-squares-fitting model, IDEAL1, and MULTI/SELF, a modified Gauss–Newton, four-exponent, self-association model included in the Beckman data analysis software. The rotor speed was 30,000 r.p.m. and equilibrium solute distributions, achieved within 18 h at 20 °C, were recorded for seven solute concentrations in the range 0.02–1.0 mg ml^{-1}, with scanning ultraviolet optics at wavelengths of 220 and 278 nm.

back of the pocket is a partially buried arginine, R471, from the β1–β2 loop, which is held in a rigid conformation by hydrogen bonds to three main-chain carbonyl oxygens, and is reminiscent of the buried arginine involved in phosphotyrosine binding to SH2 domains[9]. A positively charged residue (arginine or lysine) is present in this position in all PDZ domains that are known to bind C-terminal peptides (and in ~85% of known PDZ domains). We modelled a consensus tripeptide, TDV, into the putative binding site, to see if the peptide could be accommodated into this cavity with good stereochemistry and complementarity of non-covalent interactions (Fig. 3*d*). The lowest-energy model meets these criteria, and orients the valine side chain into the hydrophobic pocket, with the terminal carboxylate salt-bridging to the arginine. The aspartic acid side chain points out into solution; the thrconine hydroxyl makes a hydrogen bond with the carbonyl group of a conserved glycine (G477). The shortest peptides that bind PDZ domains are nine residues long; residues upstream of the C-terminal tripeptide could be accommodated within the groove that extends over the top of the domain. The modelling therefore shows that the cavity is a good candidate for the binding site, but crystal structures of PDZ–peptide complexes are required to define the details of the protein–peptide interactions and the molecular basis of specificity.

Homotypic dimer formation between PDZ domains plays a role in the localization of nNOS[4], and it has been suggested that self-association of PSD-95 proteins is necessary for receptor clustering[10]. In the crystal, the domain forms a dimer about a crystallographic two-fold axis, burying 13% of the monomer surface (total buried surface, 1,440 Å^2). A ridge formed by the exposed residues in strands β2 and β3 contacts the surface surrounding the putative peptide-binding pocket in a second molecule. To investigate a possible functional role for the PDZ-3 dimer, we measured its dissociation constant in solution by analytical ultracentrifugation (Table 1). The value obtained, $K_d = 2 \pm 1$ mM, shows that the dimer interaction is weak. However, at sites of receptor and ion-channel clustering where the local concentration of PDZ-containing proteins is very high, it is possible that self-association through this domain is important physiologically. □

Received 7 May; accepted 15 July 1996.

1. Ponting, C. P. & Phillips, C. *Trends Biochem. Sci.* **20,** 102–103 (1995).
2. Kim, E., Niethammer, M., Rothschild, A., Jan, Y. N. & Sheng, M. *Nature* **378,** 85–88 (1995).
3. Komau, H. C., Schenker, L. T., Kennedy, M. B. & Seeburg, P. H. *Science* **269,** 1737–1740 (1995).
4. Matsumine, A. *et al. Science* **272,** 1020–1023 (1996).
5. Brenman, J. E. *et al. Cell* **84,** 757–767 (1996).
6. Kennedy, M. B. *Trends Biochem. Sci.* **20,** 350 (1995).
7. Sato, T., Irie, S., Kitada, S. & Reed, J. C. *Science* **268,** 411–415 (1995).
8. Laskowski, R. A. *J. Mol. Graph.* **13,** 323–330 (1995).
9. Waksman, G. *et al. Nature* **358,** 646–653 (1992).
10. Gomperts, S. N. *Cell* **84,** 659–662 (1996).
11. Lue, R. A., Marfatia, S. M., Branton, D. & Chishti, A. H. *Proc. Natl Acad. Sci. USA* **91,** 9818–9822 (1994).
12. Otwinowski, Z. & Minor, W. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, 1993).
13. Knight, S. thesis, Swedish Univ. Agric. Sci. Upsalla (1989).
14. Cowtan, K. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 34–38 (1994).
15. CCP4 *Acta Crystallogr.* **D50,** 760–763 (1994).
16. Terwilliger, T. & Eisenberg, D. *Acta Crystallogr.* **A39,** 813–817 (1983).
17. Jones, T. A. *Methods Enzymol.* **115,** 157–171 (1985).
18. Brünger, A. *XPLOR Version 3.1: A system for X-Ray Crystallography and NMR* (Yale University Press, New Haven, Connecticut, 1992).
19. Holm, L. & Sander, C. *J. Mol. Biol.* **233,** 123–138 (1993).
20. Jabri, E., Carr, M. B., Hausinger, R. P. & Karplus, P. A. *Science* **268,** 998–1004 (1995).
21. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. *J. Mol. Biol.* **247,** 536–40 (1995).
22. Cedergren-Zeppezauer, E., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. *Nature* **355,** 740–743 (1992).
23. Kraulis, P. J. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
24. Bacon, D. J. & Anderson, W. F. *J. Mol. Graph.* **6,** 219–220 (1988).
25. Merritt, E. A. & Murphy, M. E. P. *Acta Crystallogr.* **D50,** 869–873 (1994).
26. Nicholls, A., Sharp, K. A. & Honig, B. *Proteins Struct. Funct. Genet.* **11,** 281–296 (1991).
27. Goodford, P. J. *J. Med. Chem.* **28,** 849–857 (1985).
28. Sali, A. & Blundell, T. L. *J. Mol. Biol.* **234,** 779–815 (1993).

ACKNOWLEDGEMENTS. We thank G. Maalouf for suggesting this project, L. A. Bankston for critical reading of the manuscript, N. Errington for ultracentrifugation, P. C. E. Moody and J. White for setting up the X-ray and computer facilities, and W. V. Shaw for getting us here. M. T. is a Royal Society University Research Fellow. A.H.C. is an American Heart Association Investigator. The work was supported by the BBSRC, EPSRC and NIH.

CORRESPONDENCE and requests for materials should be addressed to R.C.L. (e-mail: rcl6@leicester.ac.uk).